# The Minimum Cost *D*-Geodiverse Anycast Routing with Optimal Selection of Anycast Nodes

Amaro de Sousa
Instituto de Telecomunicações
DETI, Universidade de Aveiro, Portugal
asou@ua.pt

Dorabella Santos
INESC Coimbra
Coimbra, Portugal
dsantos@inescc.pt

***Abstract*—Consider a geographical network with associated link costs. In anycast routing, network nodes are partitioned into two sets – the source nodes and the anycast (destination) nodes – and the traffic of each source node is routed towards the anycast node providing the minimum routing cost path. By considering a given geographical distance parameter *D*, we define an anycast routing solution as *D*-geodiverse when for each source node there are two routing paths, each one towards a different anycast node, such that the geographical distance between the two paths is at least *D*. Such a solution has the property that any disaster with a coverage diameter below *D* affecting one routing path (but without involving neither the source node nor its entire set of outgoing links) cannot affect the other path, enhancing in this way the network robustness to natural disasters. The selection of the anycast nodes has an impact both on the feasibility and cost of a *D*-geodiverse anycast routing solution. Therefore, for a desired number of anycast nodes *R*, we define the minimum cost *D*-geodiverse anycast problem (MCD-GAP) aiming to identify a set of *R* anycast nodes that obtain a minimum cost routing solution. The problem is defined based on integer linear programming and is extended to consider the existence of vulnerability regions in the network, i.e., by imposing the geographical distance *D* only between network elements belonging to the same region. We present computational results showing the tradeoff between *D* and *R* in the optimal solutions obtained with and without vulnerability regions.**

***Keywords—minimum cost routing, anycast routing, path geodiversity, integer linear programming***

## I. INTRODUCTION

Due to the rising risk of natural disasters (hurricanes, floods, earthquakes), disaster based failures are frequently disrupting telecommunication networks [1]. When a disaster occurs, the typical assumption is that all network elements covered by the disaster are shut down. It is then important to minimize the disaster impact outside the disaster area (the pre-disaster problem) and network planning and management tools, as in [2], are important for mitigating the impact of disasters, and more generally of spatially correlated failures [3]. The network preparedness to natural disasters can be improved through path geodiversity [4], i.e., to take into consideration the geographical diversity of the network topology when making routing decisions. In [5–6], a routing protocol is proposed able to provide multiple geographically diverse paths to end nodes. In [7–8], a path geodiversity strategy is exploited assuming that a pair of routing paths is defined for each pair of nodes, where the two paths are geographically separated by a minimum distance *D*. The separation is defined as the minimum distance between intermediate nodes of one path and of the other path [7] or between intermediate elements (nodes and links) of one path and all elements of the other path [8]. The aim is that a disaster with a geographical coverage lower than *D* affecting intermediate elements of one path does not affect the other path. In [9–10], two close problems are addressed both considering a generalization of the min-cut and max-flow problems under geographic failures which are modelled as circular disks. Path geodiversity has also been exploited recently in the context of optical core networks [11–13].

These works, though, consider geodiversity routing for unicast communications. On the other hand, anycast routing is becoming increasingly important. Two major examples are Content Delivery Networking (CDN) and Software Defined Networking (SDN). In CDN, content is replicated over multiple data centers (DCs) and users retrieve content from the closest DC hosting it. The placement of content replicas on a supporting network is known as the Replica Placement Problem (RPP). In SDN, the network control plane is separated from the data plane and is based on a set of physically distributed SDN controllers. Then, in a logically centralized control plane, switches query the closest (primary) controller for routing decisions. The placement of controllers over the data plane network is known as the Controller Placement Problem (CPP).

Both CDN and SDN robustness against failures has been recently addressed. In [14], integer linear programming (ILP) and heuristic methods are proposed for the combined RPP and routing assignment in optical networks, where disasters are modeled as Shared Risk Groups (SRGs). In [15], a probability disaster model is assumed and a disaster-aware dynamic content-management algorithm is proposed to dynamically adapt the replica placement according to disaster probability updates so that the expected content loss is reduced at any time (see [16] for a more detailed survey). In [17–19], the CDN robustness to link cut attacks is studied which includes appropriate vulnerability measures [17–18] and a method to determine RPP Pareto-optimal solutions between user-to-replica distance and attack robustness [18]. Different works have also addressed resilient CPP variants in the context of SDN. In [20], controllers are assumed to fail with a given probability and the average delays take into account failure probabilities. In [21], a CPP is proposed guaranteeing two node disjoint paths from each switch to its primary controller, and another CPP guaranteeing node disjoint paths from each switch to its primary and to its

First version of the paper was submitted while Dorabella Santos was with Instituto de Telecomunicações, Aveiro, Portugal. During the final version revision, she was with INESC-Coimbra which she joined on 1 Dec. 2018.

backup controller. In [22], the resilient capacitated CPP is addressed considering multiple controller failures where each switch has a given traffic load and controllers have an associated capacity. Assuming that an attacker knows the data plane topology but is unaware of the controller locations, the network vulnerabilities to centrality-based attacks are studied in [23] and the controller locations are proposed to be the nodes least chosen by the different attacks. In [24], CPP solutions are based on a failure correlation assessment of nodes (and links) and different types of minimal cut sets (composed of nodes and/or links) are considered to assess the network unavailability. Optimization methods are proposed in [25–26] to compute the most robust CPP solutions to malicious node attacks.

None of these works exploit geodiversity routing as a means to enhance the network preparedness to natural disasters. The closest work is [14] which studies the content placement, routing, and protection of paths on a datacenter over optical network scenario but SRGs modeling disasters must be known a priori. In SDN, [21] exploits the idea of having two node disjoint paths from each switch (one to its primary and another to its backup controller) to protect the SDN control plane to single node failures and geodiversity routing is a natural extension of such work. To the best of our knowledge, the selection of the anycast nodes exploiting the path geodiversity concept has not been addressed.

Consider a given network with associated link costs. In anycast routing, network nodes are partitioned into two sets – the source nodes and the anycast nodes – and the traffic of each source node is routed towards the anycast node providing the minimum routing cost path. For a given geographical distance $D$, we define an anycast routing solution as being $D$-geodiverse when for each source node there are two routing paths, each one towards a different anycast node, such that the geographical distance between the two paths is at least $D$. Similar to the unicast case, this solution has the property that any disaster with a coverage diameter below $D$ affecting one routing path (but without involving neither the source node nor its entire set of outgoing links) cannot affect the other path. Then, for a desired number of anycast nodes $R$, we define the minimum cost $D$-geodiverse anycast problem (MCD-GAP) aiming to identify a set of $R$ anycast nodes that obtain a minimum cost routing solution. MCD-GAP is defined based on integer linear programming and is extended to consider the existence of vulnerability regions in the network, i.e., by imposing the minimum geographical distance $D$ only between network elements belonging to the same region.

This paper is organized as follows. Section II defines the concept of a pair of $D$-geodiverse paths in the anycast context. Section III describes the MCD-GAP through integer linear programming both with and without vulnerability regions. Section IV presents the computational results and analyses the tradeoff between the geographical distance $D$ and the number of anycast nodes $R$ in the optimal solutions. Finally, Section V presents the main conclusions of the work.

## II. Pair of $D$-Geodiverse Anycast Paths

For a given set of anycast nodes on a given geographical network, a pair of $D$-geodiverse anycast paths starting at a given source node is a pair of node disjoint routing paths, each one towards a different anycast node, and whose minimum distance is at least $D$. Consider the example of Fig. 1 with 8 nodes (1 to 8) and 11 links ($a$ to $k$) whose geographical routes are represented as graph edges. In this example, consider the source node 1 and the anycast nodes 3, 5 and 8 (highlighted in gray).

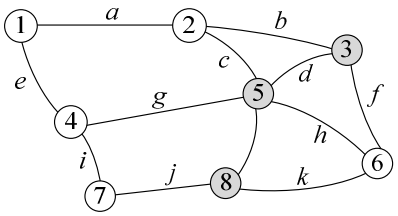



Fig. 1. Geographical network example.

Fig. 2 presents 3 pairs of anycast paths (thick links) with the minimum distance between each pair highlighted in red. In Fig. 2a, the minimum distance $D_1$ is the minimum distance between node 5 and link $b$. In this case, any disaster with a geographical coverage below $D_1$ can only affect both paths if it covers either the source node 1 or simultaneously links $a$ and $e$. Otherwise, source node 1 will always be able to communicate with an anycast node. In Fig. 2b, the minimum distance $D_2$ is the distance between nodes 5 and 8 and in Fig. 2c, the minimum distance $D_3$ is the minimum distance between node 4 and link $a$. Assuming $D_1 < D_2 < D_3$, if the required $D$ is lower than $D_1$, all solutions are feasible and the best solution is the one whose total link cost in both paths is minimal. Otherwise, the feasible solutions are the ones whose minimum distance is not lower than $D$. If $D > D_3$, none of the solutions is feasible for these anycast nodes. In this case, either a different set of anycast nodes can enable the existence of a pair of $D$-geodiverse anycast paths for node 1 or a feasible solution is always to consider node 1 as an anycast node.

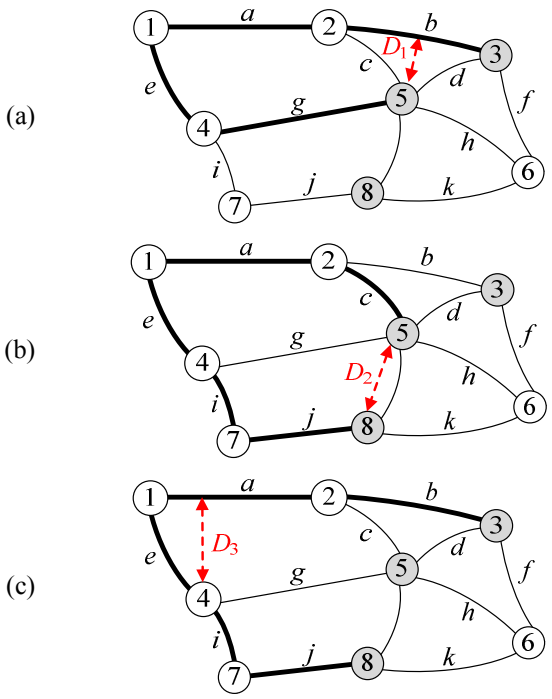



Fig. 2. Three pairs of anycast paths for source node 1.

The geographical distance between two links $a$ and $b$, denoted by $\delta(a,b)$, is the minimum distance between any point of one link and any point of the other link. Since the geographical distance between two links $a$ and $b$ is always

lower or equal to the geographical distance between any end node of $a$ and any end node of $b$, to guarantee the minimum distance $D$ between two routing paths, we only need to consider the distances between links.

Nevertheless, special care is required for the links incident to the source node. In a pair of routing paths, one of these links must belong to one path and another of these links must belong to the other path and, by the previous definition, the distance between these links is 0 (they share the source node). In this case, we define their distance based on their non-common end nodes. Consider two links $a$ and $b$ sharing source node $s$ whose end nodes are $(s,i)$ and $(s,j)$, respectively. Consider $\gamma_a = \delta(s,i)$ and $\gamma_b = \delta(s,j)$ as the distance between the end nodes of links $a$ and $b$, respectively. The distance $\beta_{ab}$ from $a$ to $b$ is defined as $\beta_{ab} = \delta(i,b)$ if $\delta(i,b) < \gamma_a$, or $\beta_{ab} = +\infty$ otherwise. Similarly, the distance $\beta_{ba}$ from $b$ to $a$ is defined as $\beta_{ba} = \delta(j,a)$ if $\delta(j,a) < \gamma_b$, or $\beta_{ba} = +\infty$ otherwise. Then, the geographical distance $\delta(a,b)$ is defined as $\delta(a,b) = \min(\beta_{ab}, \beta_{ba})$.

Fig. 3a shows one example of a source node $s$ connected to three neighbor nodes: node 1 through link $a$, node 2 through link $b$ and node 3 through link $c$. Consider first the pair of links $a$ and $b$. In Fig. 3b, the minimum distance $\beta_{ab}$ from node 1 (the non-common end node of link $a$) to link $b$ is shorter than the distance $\gamma_a$ from node 1 to the source node $s$. On the other hand, the minimum distance from node 2 to link $a$ is $\gamma_b$. So, $\delta(a,b) = \min(\beta_{ab}, +\infty) = \beta_{ab}$. In the other two link pairs ($a$ and $c$, $b$ and $c$), the minimum distance between the non-common end node of one link and the other link is always its distance to the source node and, so, $\delta(a,c) = \delta(b,c) = +\infty$. In practice, considering these distances infinite is equivalent to ignoring their distance. The reason is that if both links are in the two routing paths, the minimum coverage disaster that shuts down the non-common end node of one link and the other link also shuts down the source node and such a disaster cannot be protected (recall that a disaster involving the source node cannot be protected by any pair of anycast paths).

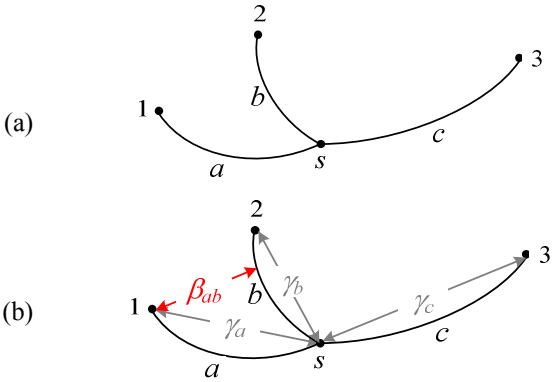



Fig. 3. Geographical distance betweek links incident to the source node $s$.

Consider a geographical network with a set of nodes $N$ and a set of links $L$. Since the geographical distances, as defined before, depend on the source node $s \in N$, we define set $P_s$ for a given minimum distance $D$ and for each node $s \in N$ composed by the pairs of links $p$ and $q$ such that $\delta(p,q) < D$, i.e., $P_s$ is the set of link pairs that cannot belong simultaneously to both routing paths (one in each path). The different sets $P_s$ differ between them only on the pairs of links sharing the source node $s$. These sets will be used in next section to define the appropriate geodiversity constraints. Note that, by default, each set $P_s$ includes all pairs of links with a common end node when this node is not the source node $s$ (their geographical distance is 0, which is always below $D$). For reasons that will be clear in the next section, we exclude these link pairs from each set $P_s$.

## III. Minimum Cost $D$-Geodiverse Anycast Problem

Consider the set of nodes $N$ and the set of links $L$ of a geographical network. Each link $p \in L$ connecting two nodes is modelled by two opposite arcs $(p_1, p_2)$ and $(p_2, p_1)$ where $p_1$ and $p_2$ are the link end nodes. Consider the network modelled by a directed graph $G = (N, A)$ where $A$ is set of all arcs of the network and $c_{ij}$ is the cost of using arc $(i,j) \in A$ on each routing path. Arc costs can have different meanings but in most CDN and SDN works represent delays so that the routing cost minimization aims to minimize the average routing delay. For a desired geographical distance $D$ and number of anycast nodes $R$, the MCD-GAP aims to select $R$ anycast nodes such that for each other node in the network there is a pair of $D$-geodiverse anycast paths and the total routing cost is minimized.

### A. MCD-GAP without Vulnerability Regions

For each node $s \in N$ and the minimum geographical distance $D$, consider the set of link pairs $P_s$ as defined in Section II. MCD-GAP is formulated as an integer linear programming (ILP) model with the following variables:

| | |
|---|---|
| $r_i \in \{0,1\}$ | binary variable that is 1 if node $i$ is selected as an anycast node, or 0 otherwise |
| $a_i^s \in \{0,1\}$ | binary variable that is 1 if node $i$ is the destination node of the first path of source node $s$, or 0 otherwise |
| $b_i^s \in \{0,1\}$ | binary variable that is 1 if node $i$ is the destination node of the second path of source node $s$, or 0 otherwise |
| $x_{ij}^s \in \{0,1\}$ | binary variable that is 1 if arc $(i,j)$ belongs to the first path of source node $s$, or 0 otherwise |
| $y_{ij}^s \in \{0,1\}$ | binary variable that is 1 if arc $(i,j)$ belongs to the second path of source node $s$, or 0 otherwise |

Consider $V(i)$ as the set of all neighboring nodes of $i$, i.e., all nodes that are connected to $i$ by an arc in $A$. Also consider the binary parameters $t_i^s$, for all $s, i \in N$, which are equal to 1 when $i = s$ and equal to 0 when $i \neq s$. An ILP model defining MCD-GAP is given by:

$$\text{Min} \sum_{s \in N} \sum_{(i,j) \in A} c_{ij}(x_{ij}^s + y_{ij}^s) \quad (1)$$

Subject to:

$$\sum_{i \in N} r_i = R \quad (2)$$

$$\sum_{j \in V(i)} (x_{ji}^s - x_{ij}^s) = a_i^s - t_i^s \qquad s \in N, i \in N \quad (3)$$

$$\sum_{j \in V(i)} (y_{ji}^s - y_{ij}^s) = b_i^s - t_i^s \qquad s \in N, i \in N \quad (4)$$

$$\sum_{j \in V(i)} (x_{ji}^s + y_{ji}^s) \leq 1 \qquad s \in N, i \in N \backslash \{s\} \quad (5)$$

$$a_i^s + b_i^s \leq r_i \qquad s \in N, i \in N \backslash \{s\} \quad (6)$$

$$a_s^s + b_s^s = 2r_s \qquad s \in N \quad (7)$$

$$\sum_{i \in N} (a_i^s + b_i^s) = 2 \qquad s \in N \quad (8)$$

$$x^s_{p_1p_2} + x^s_{p_2p_1} + y^s_{q_1q_2} + y^s_{q_2q_1} \leq 1 \qquad s \in N, (p,q) \in P_s \quad (9)$$

$$x^s_{q_1q_2} + x^s_{q_2q_1} + y^s_{p_1p_2} + y^s_{p_2p_1} \leq 1 \qquad s \in N, (p,q) \in P_s \quad (10)$$

$$x^s_{ij}, y^s_{ij} \in \{0,1\} \qquad s \in N, (i,j) \in A \quad (11)$$

$$a^s_i, b^s_i \in \{0,1\} \qquad s \in N, i \in N \quad (12)$$

$$r_i \in \{0,1\} \qquad i \in N \quad (13)$$

The objective function (1) is the minimization of the total cost of all pairs of routing paths. Constraints (2) guarantee that a total of $R$ nodes are selected as anycast nodes.

Constraints (3–4) are the path conservation constraints for the two routing paths of each source node $s \in N$. Constraints (3) guarantee that variables $x^s_{ij}$ define the first routing path from node *s* to the destination node *i* (i.e., the node *i* such that $a^s_i = 1$). Constraints (4) are the equivalent to constraints (3) for the second routing path from node *s* to the node *i* such that $b^s_i = 1$. If both destination nodes are the source node *s* (i.e., when $a^s_s = 1$ and $b^s_s = 1$) constraints (3) let variables $x^s_{ij}$ be set to 0 and constraints (4) let variables $y^s_{ij}$ be set to 0, which means that constraints (3–4) let both routing paths be empty. Constraints (5) guarantee that the two routing paths cannot share any node besides the source node *s* (i.e., the two routing paths are node disjoint).

Constraints (6–8) guarantee a proper selection of destination nodes for all pairs of routing paths. Constraints (6) guarantee that all destination nodes are anycast nodes. If the source node *s* is an anycast node, constraints (7) guarantee that it is also the destination node of its two routing paths (recall that in this case, constraints (3–4) allow the routing paths to be empty). Otherwise, constraints (8), together with constraints (6), guarantee that the destination nodes are two different anycast nodes for each source node *s*.

Constraints (9–10) are the geodiversity constraints, i.e., they guarantee that all pairs of routing paths are *D*-geodiverse. For each node $s \in N$ and each pair of links $(p,q) \in P_s$, if one of the arcs of link $p$ is in the first path (constraints (9)) or in the second path (constraints (10)), none of the arcs of link $q$ can be in the second path nor in the first path, respectively. Otherwise, if one of the arcs of link $q$ is in the first path (constraints (10)) or in the second path (constraints (9)), none of the arcs of link $p$ can be in the second path nor on the first path, respectively. Finally, constraints (11–13) are variable domain constraints.

Depending on value *D*, the number of link pairs of sets $P_s$ (and, consequently, the number of constraints (9–10)) can be too large. The node disjoint constraints (5) are a compact set of constraints guaranteeing that the pairs of links sharing a common end node cannot be simultaneously in the pair of routing paths. Therefore, such link pairs do not need to be included in sets $P_s$ (as described at the end of Section II). Our computational tests showed that this approach not only reduces the solution runtime of many problem instances but also reduces the number of instances that cannot be solved to optimality due to out-of-memory issues.

Nevertheless, this ILP model has the symmetry problem that makes it hard to be solved for all our instances of interest. This problem is due to the fact that each pair of routing paths is represented by two mathematical solutions, one solution where variables $x^s_{ij}$ and $y^s_{ij}$ represent the paths in one order and another where the same variables represent the paths in the reverse order. We mitigated the symmetry problem with the following variable elimination rule. For each source node $s \in N$, we take the first pair of links $(p,q) \in P_s$ with *s* as a common end node. If such link pair does not exist, we do nothing. If it exists, since both links cannot be simultaneously used, we impose (without any lack of generality) that they can be used only in the first routing path of source node *s* by setting to 0 the four variables ($y^s_{p_1p_2}$, $y^s_{p_2p_1}$, $y^s_{q_1q_2}$ and $y^s_{q_2q_1}$) associated with the second routing path. Our computational tests showed that this variable elimination rule further reduces the solution runtime of many problem instances and solves the out-of-memory issues of almost all of the problem instances.

### *B. MCD-GAP with Vulnerability Regions*

MCD-GAP, as defined by the previous ILP model, assumes that a disaster can happen at any geographical area of the network and, therefore, requires all pairs of anycast paths to be geographically separated by at least *D* in all their extension. In practice, the probability of natural hazards is not uniform and network operators might want to tailor the network robustness to the different hazard types and regions of their networks, which are referred as vulnerability regions. Moreover, the network operator might consider a different minimum geographical separation between each pair of paths for each vulnerability region, depending on its hazard type. So, consider a set of *V* vulnerability regions and a minimum geographical distance $D_v$ associated to each region $v = 1 \ldots V$. The aim is that each pair of paths must be node disjoint outside regions and must be $D_v$-geodiverse inside each vulnerability region $v = 1 \ldots V$.

The proposed ILP model can consider vulnerability regions by only redefining sets $P_s$. We first compute the links belonging to each vulnerability region as the ones that are incident on nodes belonging to the region and/or intersect the region (each link can belong to multiple regions). Then, for each node $s \in N$, the set $P_s$ is composed by all pairs of links $(p,q)$ that belong to the same region *v* such that $\delta(p,q) < D_v$. If a pair of links belongs to more than one vulnerability region, we impose the largest minimum geographical distance among all involved regions.

## IV. COMPUTATIONAL RESULTS

We have conducted a set of computational tests in order to assess the efficiency of branch-and-cut methods to solve the MCD-GAP and to analyze the tradeoff between the geographical distance *D* and the number of anycast nodes *R* in the optimal solutions. The ILP model was implemented in C++ using CPLEX 12.6 callable libraries running 8 threads. All problem instances were solved on a 16 core server with 64 GB of RAM running Windows OS. In the tests, we have used Germany50 network (Fig. 4a) with 50 nodes, 88 links and an average node degree of 3.52 (information available at sndlib.zib.de) and CORONET CONUS network (Fig. 5a) with 75 nodes, 99 links and an average node degree of 2.64 (information available at monarchna.com/topology.html). In both cases, all distances were computed considering that links follow the shortest path over the terrestrial surface. With these assumptions, the graph diameter (the longest shortest path between any two nodes) is 934 km for Germany50 and 6472 km for CORONET CONUS.

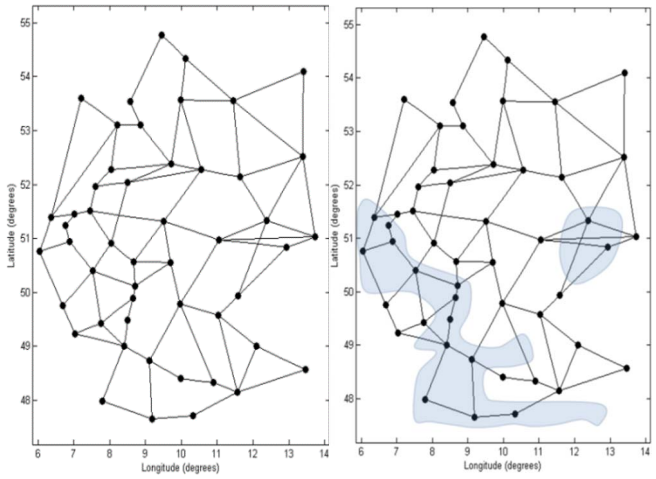



Fig. 4. Germany50 without (a) and with (b) vulnerability regions.

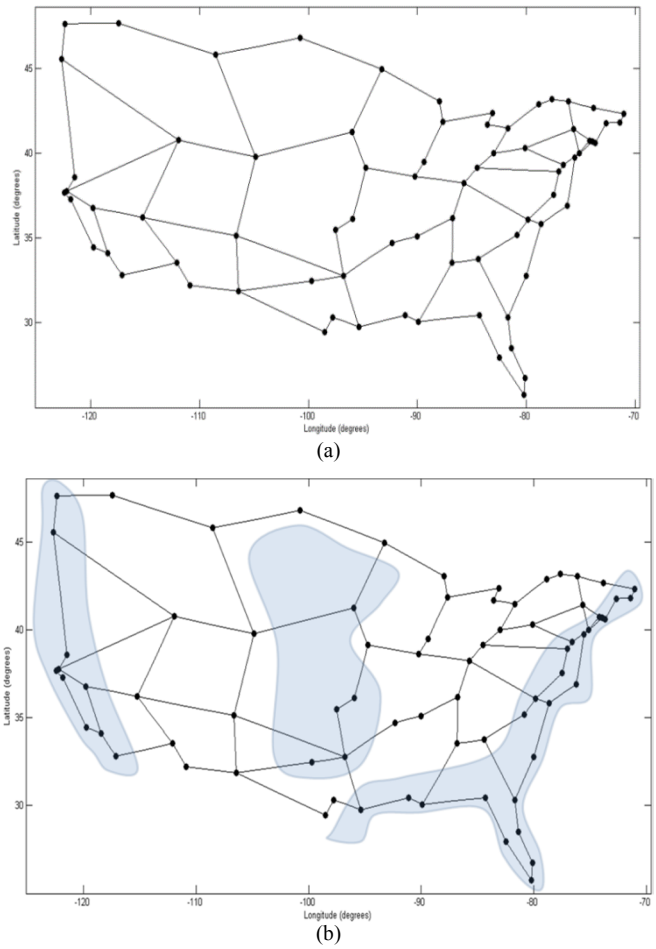



Fig. 5. CORONET CONUS without (a) and with (b) vulnerability regions.

For the instances with vulnerability regions, we have considered publicly available natural hazard maps for the countries harboring the two networks. For Germany50, we considered the Germany seismic hazard map [27]. An approximation of the two most hazardous regions (level VI-VIII in [27]) was used to define the two vulnerability regions shown in Fig. 4b. For CORONET CONUS, we considered the USA natural hazard risk map which jointly considers risk maps for earthquakes, floods, tornados and hurricanes (available at http://alertsystemsgroup.com/earthquake-early-warning/informative-maps). Approximations of the most hazardous regions were considered to define the three vulnerability regions shown in Fig. 5b: the left region is most prone to earthquakes, the middle region is most prone to tornados and the right region is most prone to hurricanes.

In the computational tests, we have assumed five different values for the geographical distance: $D$ = 20, 40, 60, 80 and 100 Km for the smallest Germany50 and $D$ = 60, 120, 180, 240 and 300 Km for the largest CORONET CONUS. For comparative analysis, the instances with vulnerability regions consider the same value of $D$ in all regions (i.e., $D_v = D$ for all $v = 1 \ldots V$). For each value of $D$, we have solved the problem instance for all meaningful values of $R$, i.e., for $R = 2 \ldots |N| - 1$.

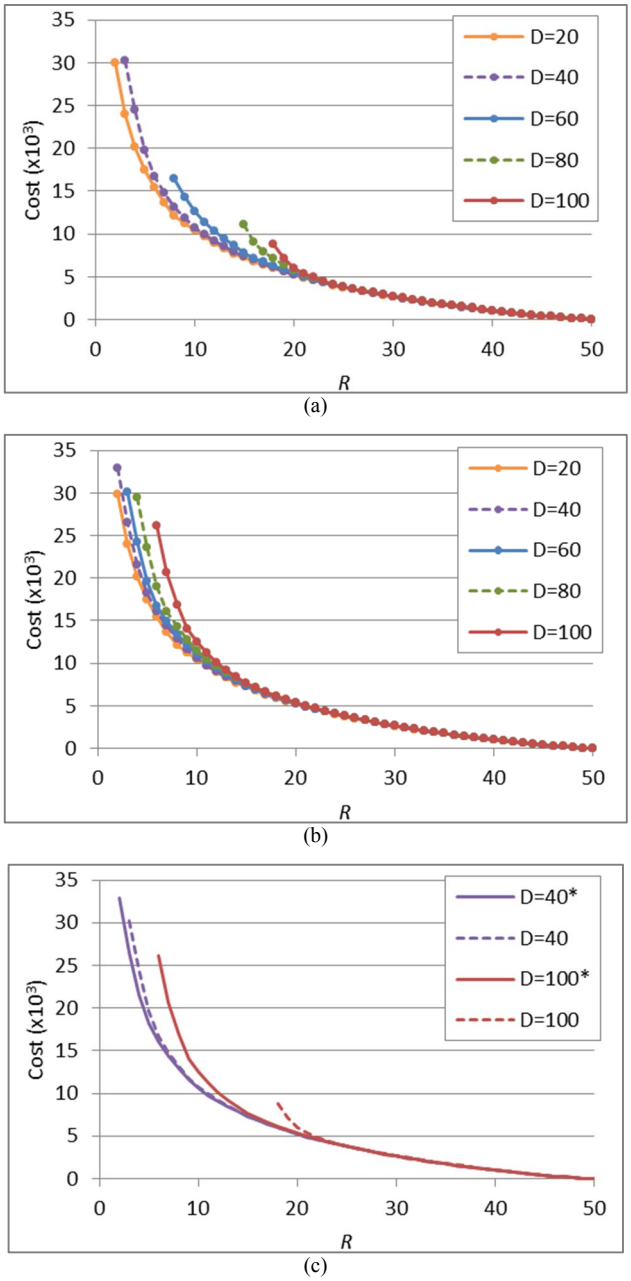



Fig. 6. Optimal costs of Germany50 without (a) and with (b) vulnerability regions. In (c), the costs with (labelled with *) and without vulnerability regions for $D$ = 40 and 100 Km.

Fig. 6 presents the optimal cost values of all Germany50 feasible problem instances without vulnerability regions (Fig. 6a) and with vulnerability regions (Fig. 6b). For comparison

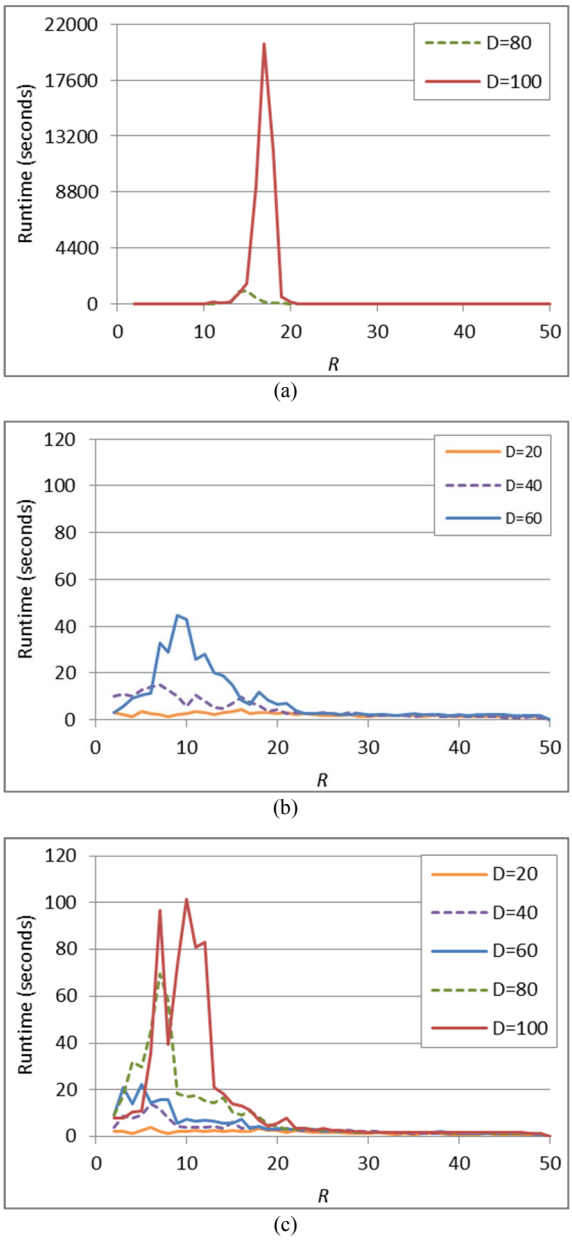
reasons, we repeat in Fig. 6c the results with and without vulnerability regions for $D$ = 40 and 100 Km (the pairs of values $D$ and $R$ not present in these figures are the ones whose problem instances are infeasible). The results of Fig. 6a and 6b show some expected conclusions. Firstly, the optimal solution cost increases with larger geographical distances $D$: when requiring more geographically separated pairs of anycast paths, they become longer and, therefore, the average cost of all pairs of anycast paths becomes higher. Secondly, the optimal solution cost decreases with higher number of anycast nodes $R$: with more anycast nodes, there are more possible destination nodes nearby each source node and, therefore, the average cost of all pairs of anycast paths becomes lower. Finally, the minimum number of anycast nodes $R$ increases with larger values of $D$: when requiring more geographically separated pairs of anycast paths, more nodes must be selected as anycast nodes (recall the discussion in Section II with the example of Fig. 2).

Comparing the results with and without vulnerability regions, we reach two key conclusions. Firstly, for the pairs of values $D$ and $R$ such that both cases are feasible, there are cost gains in considering vulnerability regions. Moreover, these cost gains are higher for higher values of $D$ (in Fig. 6c, the cost differences is small for $D$ = 40 Km and become very large for $D$ = 100 Km). Secondly, the minimum required number of anycast nodes $R$ is lower for each value of $D$ when vulnerability regions are considered (i.e., we obtain the same robustness level to natural disasters with fewer anycast nodes). Moreover, this reduction is higher for higher values of $D$ (in Fig. 6c, the minimum number of anycast nodes reduces from 3 to 2 when $D$ = 40 Km and 18 to 6 when $D$ = 100 Km). So, the more robust against natural disasters the operator aims for its network, the more important it is to consider vulnerability regions. A careful characterization of the vulnerability regions lets the operator get either more robust solutions for the same number of anycast nodes or lower number of anycast nodes to reach the same desired robustness to natural disasters.

Concerning the efficiency of CPLEX while solving the Germany50 instances, our tests have shown that the 32-bit CPLEX version is, on average, more efficient than the 64-bit CPLEX version, as long as the required memory does not explode. By default, we have solved all instances with the 32-bit version. Six of all instances ended by out-of-memory: the instances without vulnerability regions for $D$ = 100 Km and $R$ from 14 to 19. These instances were then solved successfully with the 64-bit version. Fig. 7 presents the CPLEX runtime values (in seconds) of all Germany50 instances. The instances without geographical regions are separated in: the hardest cases of $D$ = 80 and 100 Km (Fig. 7a) and the easiest cases of $D$ = 20, 40 and 60 Km (Fig. 7b). The instances with geographical regions are presented together for all cases (Fig. 7c).

These results include the runtime values of the infeasible problem instances. These results show that, without vulnerability regions, the instances become very hard to solve for larger values of $D$. For $D$ up to 60 Km, the worst runtime is under 50 seconds (Fig. 7b) and becomes 1030 seconds for $D$ = 80 Km and almost 20500 seconds (around 5 hours and 40 minutes) for $D$ = 100 Km (Fig. 7a). On the other hand, the instances with vulnerability regions are much easier (Fig. 7c) in all cases although, once again, the runtime values are higher for larger values of $D$.

Fig. 8 presents the optimal cost values of all CORONET CONUS feasible problem instances for the cases without (Fig. 8a) and with vulnerability regions (Fig. 8b). Once again, we repeat in Fig. 8c the results with and without vulnerability regions now for $D$ = 120 and 300 Km. These results exhibit the same trends but with greater differences than the previous Germany50 results.



(a)

(b)

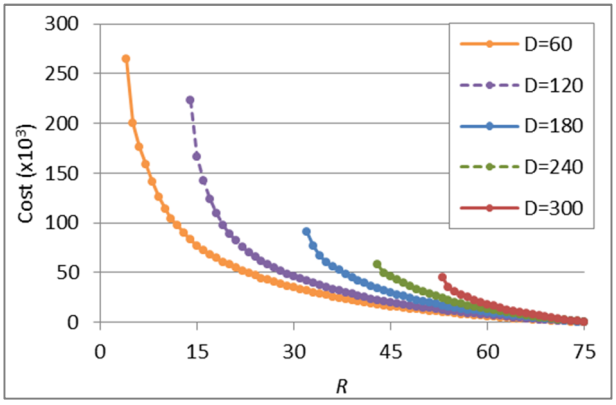
(c)

Fig. 7. CPLEX runtime of Germany50 instances without survivability regions for $D$ = 80 and 100 Km (a) and for $D$ = 20, 40 and 60 Km (b) and with survivability regions (c).

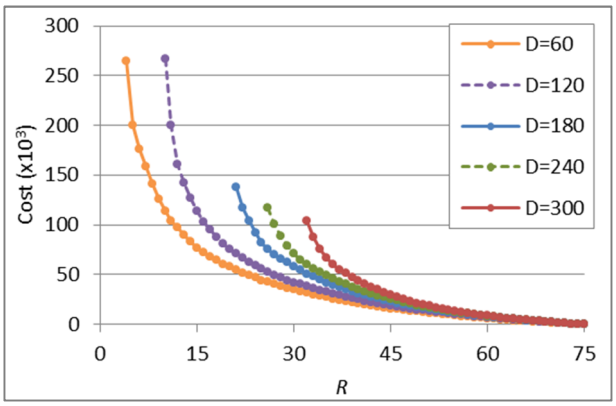
The results of Fig. 8c show that for the smaller value of $D$ = 120 Km, the minimum number of anycast nodes is $R$ = 14 with a cost of 223008 without vulnerability regions (the anycast nodes of the optimal solution shown in Fig. 9a). On the other hand, with the vulnerability regions the minimum number of anycast nodes is $R$ = 10 with a cost of 266560 while this cost is reduced to 200217 if $R$ = 11 and to 126362 if $R$ = 14 (in the latter case, the anycast nodes of the optimal solution are shown in Fig. 9b). So, by considering the vulnerability regions, the operator can obtain the same robustness against natural disasters with: (i) a cost reduction

of 10.2% (from 223008 to 200217) and at the same time a reduction on the number of anycast nodes from 14 to 11 or (ii) a cost reduction of 43.3% (from 223008 to 126362) with the same number of 14 anycast nodes.

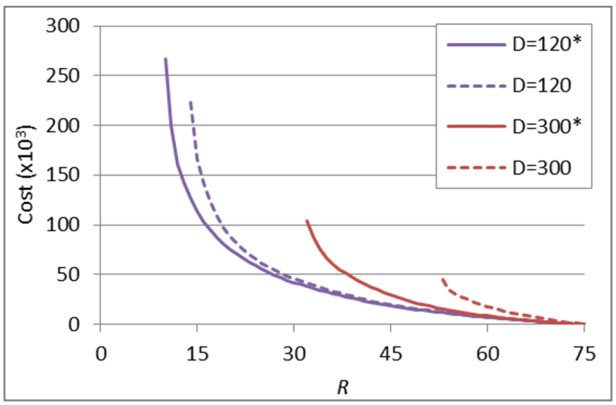
The optimal anycast nodes presented in Fig. 9 highlight a very interesting observation. To fulfil the geodiversity requirements, anycast nodes are mainly located on the network parts containing closer nodes and shorter links in the whole network (Fig. 9a) or only inside the vulnerability regions (Fig. 9b). When the vulnerability regions are considered, since outside these regions there are no geodiversity requirements, anycast nodes are placed more uniformly throughout the network obtaining in this way the huge cost reduction of 43.3% as already observed.



(a)

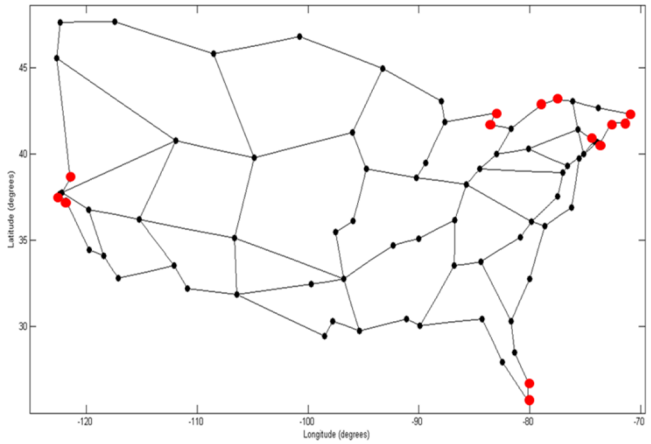


(b)

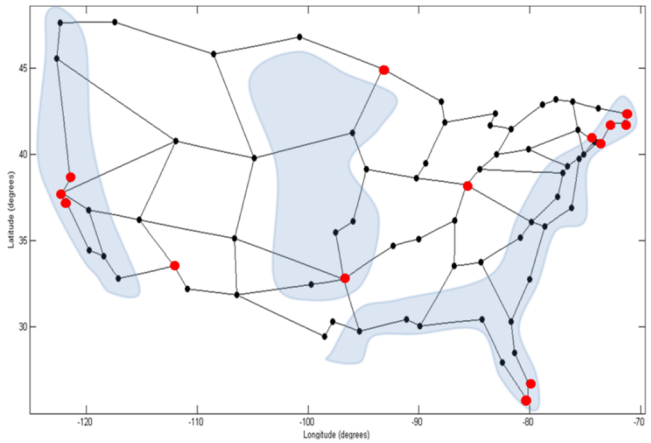


(c)

Fig. 8. Optimal costs of CORONET CONUS without (a) and with (b) vulnerability regions. In (c), the costs with (labelled with *) and without vulnerability regions for $D$ = 120 and 300 Km.

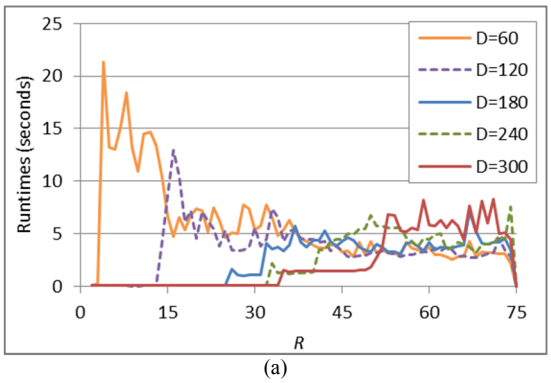


(a)

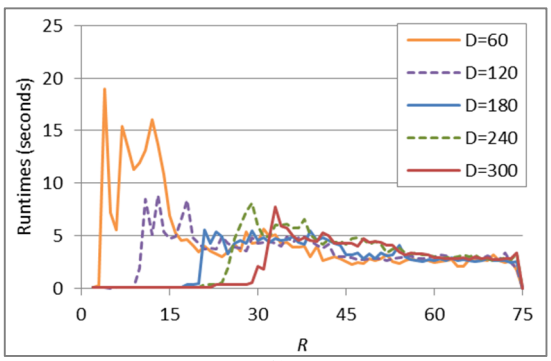


(b)

Fig. 9. Anycast nodes (in red) of CORONET CONUS for $D$ = 120 Km and $R$ = 14 anycast nodes without (a) and with (b) vulnerability regions.



(a)



(b)

Fig. 10. CPLEX runtime of CORONET CONUS without (a) and with (b) survivability regions.

Fig. 10 presents the CPLEX runtime values (in seconds) of all CORONET CONUS instances without (Fig. 10a) and with vulnerability regions (Fig. 10b). All instances were now successfully solved by the 32-bit CPLEX version. These instances are easier to solve than the Germany50 ones since the worst runtime value is below 22 (and 19) seconds among all problem instances without (and with) vulnerability regions. Note that the ILP models of CORONET CONUS instances are much larger (more variables and constraints) than the ones of Germany50 instances (usually, larger ILP models are harder to solve). In this case, the average node degree of CORONET CONUS is much lower than the one of Germany50, which implies that the number of possible paths between pairs of nodes is smaller which, in turn, makes the problems easier to solve. As a final observation, unlike Germany50 where the highest runtime values were for the larger values of $D$, for CORONET CONUS the highest runtime values are for the smallest values of $D$.

## V. Conclusions

In this work, we have exploited path geodiversity in anycast communications as a means to enhance the network robustness against natural disasters. We have defined and solved the minimum cost $D$-geodiverse anycast routing problem with optimal selection of anycast nodes, based on integer linear programming, and extended it to consider the existence of vulnerability regions. We have presented computational results based on two well-known network topologies using real information of their hazard regions.

We were able to compute the optimal solutions for all parameters of interest. The results showed that, in general, improving the robustness to natural disasters increases the routing costs and requires a higher minimum number of anycast nodes. More importantly, a careful characterization of the vulnerability regions allows the operator to achieve either improved robustness with the same cost and number of anycast nodes or reduced number of anycast nodes and routing costs for the same robustness. Finally, anycast nodes in optimal solutions are mainly located on the network parts containing closer nodes and shorter links in the whole network, if no vulnerability regions are considered, or only inside vulnerability regions, otherwise.

## Acknowledgments

This paper is based upon work from COST Action CA15127 ("Resilient communication services protecting end user applications from disaster-based failures – RECODIS") supported by COST Association. The work was financially supported by FCT, Portugal, under the projects CENTRO-01-0145-FEDER-029312 and UID/EEA/50008/2013 and through the postdoc grant SFRH/BPD/ 111503/2015.

## References

[1] J. Rak, et al., "RECODIS: Resilient communication services protecting end-user applications from disaster-based failures", ICTON, We.D1.4, 2016.

[2] A. Das, A. Sen, C. Qiao, N. Ghani, N. Mitton, "A network planning and management tool for mitigating the impact of spatially correlated failures in infrastructure networks", DRCN, pp. 71–78, 2016.

[3] T. Gomes, et al., "A survey of strategies for communication networks to protect against large-scale natural disasters", RNDM, pp. 11–22, 2016.

[4] J. P. Rohrer, A. Jabbar, and J. P. Sterbenz, "Path Diversification for Future Internet End-to-End Resilience and Survivability," Telecommunication Systems, vol. 56, pp. 49–67, May 2014.

[5] Y. Cheng, M. T. Gardner, J. Li, R. May, D. Medhi, J. P. Sterbenz, "Optimised heuristics for a geodiverse routing protocol," DRCN, pp. 1-9, 2014.

[6] Y. Cheng, M. T. Gardner, J. Li, R. May, D. Medhi, J. P. Sterbenz, "Analysing GeoPath Diversity and Improving Routing Performance in Optical Networks," Computer Networks, vol. 82, pp. 50–67, May 2015.

[7] S. Trajanovski, F. Kuipers, A. Ilic, J. Crowcroft, P. Van Mieghem, "Finding critical regions and region-disjoint paths in a network", IEEE/ACM Trans on Netwoking, vol. 23, no. 3, pp. 908-921, 2015.

[8] A. de Sousa, D. Santos, P. Monteiro, "Determination of the minimum cost pair of $D$-geodiverse paths", DRCN, pp. 101–108, 2017.

[9] Y. Kobayashi, K. Otsuki, "Max-Flow Min-Cut Theorem and Faster Algorithms in a Circular Disk Failure Model", IEEE INFOCOM, pp. 1635–1643, 2014.

[10] S. Neumayer, A. Efrat, E. Modiano, "Geographic max-flow and min-cut under a circular disk failure model", Computer Networks, vol. 77, pp. 117–127, Feb. 2015.

[11] A. de Sousa, T. Gomes, R. Girão-Silva, L. Martins, "Minimizing the network availability upgrade cost with geodiversity guarantees", RNDM, pp. 1–8, 2017.

[12] M. Ashraf, S. Idrus, F. Iqbal, R. Butt, "On spatially disjoint lightpaths in optical networks", Photonic Network Communications, vol. 23, no. 1, pp. 11–25, 2018.

[13] B. Nedic, M. Gunkel, T, Gomes, R. Girão-Silva, "SRLG-disjointness and geodiverse routing - a practical network study and operational conclusions", RNDM, pp. 1–8, 2018.

[14] M. F. Habib, M. Tornatore, M. De Leenheer, F. Dikbiyik, B. Mukherjee, "Design of disaster-resilient optical datacenter networks," J. Lightw. Technol., vol. 30, no. 16, pp. 2563–2573, 2012.

[15] S. Ferdousi, F. Dikbiyik, M. F. Habib, M. Tornatore, B. Mukherjee, "Disaster-aware datacenter placement and dynamic content management in cloud networks," IEEE/OSA J. Opt. Commun. Netw., vol. 7, no. 7, pp. 681–694, 2015.

[16] C. Colman-Meixner, C. Develder, M. Tornatore, B. Mukherjee, "A survey on resiliency techniques in cloud computing infrastructures and applications," IEEE Commun. Surveys Tuts., vol. 18, no. 3, pp. 2244–2281, 2016.

[17] C. Natalino, A. Yayimli, L. Wosinska, M. Furdek, "Content accessibility in optical cloud networks under targeted link cuts", ONDM, pp. 1–6, 2017.

[18] C. Natalino, A. Yayimli, L. Wosinska, M. Furdek, "Link addition framework for optical CDNs robust to targeted link cut attacks", RNDM, pp. 1–7, 2017.

[19] C. Natalino, A. de Sousa, L. Wosinska, M. Furdek, "On the Trade-offs between User-to-Replica Distance and CDN Robustness to Link Cut Attacks", RNDM, pp. 1–7, 2018.

[20] N. Perrot, T. Reynaud, "Optimal placement of controllers in a resilient SDN architecture", DRCN, pp. 145-151, 2016.

[21] P. Vizarreta, C. Mas Machuca, W. Kellerer, "Controller placement strategies for a resilient SDN control plane", RNDM, pp. 253–259, 2016.

[22] M. Tanha, D. Sajjadi, R. Rubyy, J. Pan, "Capacity-aware and Delay-guaranteed Resilient Controller Placement for Software-Defined WANs", IEEE Trans. on Networks and Service Management, vol. 15, no. 3, pp. 991–1005, 2018.

[23] D. F. Rueda, E. Calle, J. L. Marzo, "Improving the Robustness to Targeted Attacks in Software Defined Networks (SDN)", DRCN, pp. 78–85, 2017.

[24] G. Nencioni, B. E. Helvik, P. E. Heegaard, "Including Failure Correlation in Availability Modeling of a Software-Defined Backbone Network", IEEE Trans. on Network and Service Management, vol. 14, no. 4, pp. 1032–1045, 2017.

[25] D. Santos, A. de Sousa, C. Mas Machuca, "Robust SDN Controller Placement to Malicious Node Attacks", DRCN, co-located with ICIN, pp. 1–8, 2018.

[26] D. Santos, A. de Sousa, C. Mas Machuca, "Combined Control and Data Plane Robustness of SDN Networks against Malicious Node Attacks", CNSM, pp. 54–62, 2018.

[27] M. Müller, S. Vorogushyn, P. Maier, A. H. Thieken, T. Petrow, A. Kron, B. Büchele, J. Wächter, "CEDIM Risk Explorer – a map server solution in the project Risk Map Germany", Natural Hazards and Earth System Sciences, vol. 6, pp. 711-720, 2006.